

NOVEL SEVEN-TRANSMEMBRANE PROTEINS/G-PROTEIN COUPLED
RECEPTORS

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application Serial No.
60/182,061 filed February 11, 2000, the contents of which are herein incorporated by
5 reference.

FIELD OF THE INVENTION

The present invention relates to newly identified seven-transmembrane proteins,
10 including proteins that function as receptors belonging to the superfamily of G-protein-
coupled receptors. The invention also relates to polynucleotides encoding the seven-
transmembrane proteins/receptors. The invention further relates to methods using the
seven-transmembrane protein/receptor polypeptides and polynucleotides as a target for
diagnosis and treatment in seven-transmembrane protein/receptor-mediated and related
15 disorders. The invention further relates to drug-screening methods using the seven-
transmembrane protein/receptor polypeptides and polynucleotides to identify agonists
and antagonists for diagnosis and treatment. The invention further encompasses agonists
and antagonists based on the seven-transmembrane protein/receptor polypeptides and
polynucleotides. The invention further relates to procedures for producing the seven-
20 transmembrane/receptor polypeptides and polynucleotides.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest.* 92:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet.* 30:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet.* 26:403-424 (1992)), and nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet.* 2:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the β 2-adrenergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem.* 60:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science* 254:1024-1026 (1991); Lin *et al.*, *Science* 254:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science* 258 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science* 241:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem.* 61:1097-1129 (1992)).

There are also a small number of other proteins which present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein which has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5047-5051 (1993)). The gene frizzled (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, *fz* has not been shown to couple to G-proteins (Vinson *et al.*, *Nature* 338:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α , β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α -subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the $\beta\gamma$ -subunits. The GTP-bound form of the α -subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenylyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α -subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

Accordingly, GPCRs are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing novel seven-transmembrane proteins/GPCRs.

SUMMARY OF THE INVENTION

It is an object of the invention to identify novel seven-transmembrane proteins/GPCRs.

5 It is a further object of the invention to provide novel seven-transmembrane protein/GPCR polypeptides that are useful as reagents or targets in seven-transmembrane protein/receptor assays applicable to treatment and diagnosis of seven-transmembrane protein/GPCR-mediated disorders.

10 It is a further object of the invention to provide polynucleotides corresponding to the novel seven-transmembrane protein/GPCR receptor polypeptides that are useful as targets and reagents in seven-transmembrane protein/receptor assays applicable to treatment and diagnosis of seven-transmembrane protein/GPCR-mediated disorders and useful for producing novel seven-transmembrane protein/receptor polypeptides by recombinant methods.

15 A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the novel seven-transmembrane proteins/receptors.

A further specific object of the invention is to provide compounds that modulate expression of the seven-transmembrane proteins/receptors for treatment and diagnosis of
20 seven-transmembrane protein/GPCR- related disorders.

The invention is thus based on the identification of novel seven-transmembrane proteins/GPCRs, designated 17724, 31945, and 50288. As discussed more fully below, 17724 contains sequence homology or motifs/signatures that classify this protein in the GPCR superfamily, as a member of the rhodopsin family of G-protein coupled receptors.
25 The other members are putative GPCRs, and accordingly may be relevant to the various uses and methods involving GPCRs as disclosed herein.

The invention provides isolated 17724, 31945, and 50288 polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NOS:2, 5 and 8, respectively, or the amino acid sequence encoded by the cDNA deposited as ATCC No.
30 _____, _____, and _____, respectively, on _____ ("the deposited cDNA").

The invention also provides isolated 17724, 31945, and 50288 nucleic acid molecules having the sequence shown in SEQ ID NOS:1, 3, 4, 6, 7, or 9, respectively, or in the deposited cDNA.

5 The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NOS:2, 5 and 8 or encoded by the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NOS:2, 3, 4, 6, 7 and 9 or in the deposited cDNA.

10 The invention also provides fragments of the polypeptide shown in SEQ ID NOS:2, 5 and 8 and nucleotide sequence shown in SEQ ID NOS:1, 3, 4, 6, 7 and 9, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention further provides nucleic acid constructs comprising the nucleic acid molecules described above. In a preferred embodiment, the nucleic acid molecules of the invention are operatively linked to a regulatory sequence.

The invention also provides vectors and host cells for expressing the nucleic acid molecules and polypeptides of the invention and particularly recombinant vectors and host cells.

20 The invention also provides methods of making the vectors and host cells and methods for using them to produce the nucleic acid molecules and polypeptides of the invention.

The invention also provides antibodies or antigen-binding fragments thereof that selectively bind the polypeptides and fragments of the invention.

25 The invention also provides methods of screening for compounds that modulate expression or activity of the polypeptides or nucleic acid (RNA or DNA) of the invention.

The invention also provides a process for modulating polypeptide or nucleic acid expression or activity, especially using the screened compounds. Modulation may be used to treat conditions related to aberrant activity or expression of the polypeptides or nucleic acids of the invention.

The invention also provides assays for determining the presence or absence of and level of the polypeptides or nucleic acid molecules of the invention in a biological sample, including for disease diagnosis.

The invention also provides assays for determining the presence of a mutation in the polypeptides or nucleic acid molecules, including for disease diagnosis.

In still a further embodiment, the invention provides a computer readable means containing the nucleotide and/or amino acid sequences of the nucleic acids and polypeptides of the invention.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows a 17724 protein hydrophobicity plot. Relative hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and N glycosylation site (Ngly) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence (shown in SEQ ID NO:2) of human 17724 are indicated. Polypeptides of the invention include fragments which include: all or a part of a hydrophobic sequence (a sequence above the dashed line); or all or part of a hydrophilic fragment (a sequence below the dashed line). Other fragments include a cysteine residue or as N-glycosylation site.

Figure 2 shows an analysis of the 17724 amino acid sequence: α turn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 3 depicts an alignment of a domain of human 17724 with a consensus amino acid sequence derived from a hidden Markov model. The upper sequence is the consensus amino acid sequence (SEQ ID NO:10), while the lower amino acid sequence corresponds to amino acids to 125-374 of SEQ ID NO:2.

Figure 4 shows the expression pattern of the 17724 mRNA in various clinical lung samples.

Figure 5 shows the expression pattern of the 17724 mRNA in various clinical angiogenic samples (N = normal tissue, T = tumorous tissue).

Figure 6 shows expression of the 17724 protein in various normal human tissues and in diseased human heart tissues.

Figure 7 shows the expression pattern of the 17724 mRNA in various tissue samples

Figure 8 shows the expression pattern of the 17724 mRNA in various tissues samples. High levels of expression are found in tissues of the spinal cord, brain cortex and brain hypothalamus.

Figure 9 shows a 31945 protein hydrophobicity plot. Relative hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and N glycosylation site (Ngly) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence (show in SEQ ID NO:5) of human 31945 are indicated. Polypeptides of the invention include fragments which include: all or a part of a hydrophobic sequence (a sequence above the dashed line); or all or part of a hydrophilic fragment (a sequence below the dashed line). Other fragments include a cystein residue or a N-glycosylation site.

Figure 10 shows an analysis of the 31945 amino acid sequence: α βturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 11 shows expression of the 31945 mRNA in various normal and diseased human tissues and cells. It also includes expression in hepatocytes in culture and hepatocytes treated with TGF. Expression is shown in, among other things, mobilized peripheral blood (mPB), peripheral blood mononuclear cells (PBMC), and various T cells (Th).

Figure 12 shows a 50288 protein hydrophobicity plot. Relative hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and N glycosylation site (Ngly) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence (shown in SEQ ID NO:8) of human 50288 are indicated. Polypeptides of the invention include fragments which include: all or a part of a hydrophobic sequence (a sequence above the dashed line); or all or part of a hydrophilic fragment (a sequence below the dashed line). Other fragments include a cysteine residue or a N-glycosylation site.

Figure 13 shows an analysis of the 50288 amino acid sequence: α turn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 14 shows the expression pattern of 17724 mRNA in various tissues and cell types.

Figure 15 shows the expression pattern of 17724 mRNA in various tumorous and normal tissues and cell types.

Figure 16 shows the expression pattern of 17724 mRNA in various tissues and cell types.

Figure 17 shows the expression pattern of 17724 mRNA in various tissues.

DETAILED DESCRIPTION OF THE INVENTION

Receptor function/signal pathway

5 The 17724 receptor protein is a GPCR that participates in signaling pathways. The other seven-transmembrane proteins are putative GPCRs that participate in signaling pathways. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR. Examples of such functions include mobilization of intracellular molecules that participate
10 in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP₂), inositol 1,4,5-triphosphate (IP₃) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival. The 17724 protein is expressed in the tissues shown in Figures 4-8. Therefore, cells
15 participating in a 17724 protein signaling pathway include, but are not limited to, cells derived from these tissues, especially those tissues in which the gene is highly expressed. Since the 31945 protein is expressed in the tissues shown in Figure 11, cells participating in a 31945 protein signaling pathway include, but are not limited to, cells derived from these tissues, especially those cells or tissues in which the gene is highly expressed. Since the
20 50288 protein is expressed in adrenal gland, brain, breast, colon to liver metastases, pituitary, prostate and T-cell, cells participating in a 50288 protein signaling pathway include, but are not limited to, cells derived from this tissue.

 The response mediated by a receptor protein depends on the type of cell. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity
25 such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, it is universal that a GPCR of the invention interacts with G proteins to produce one or more secondary signals, in a

variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP₂) as well as to the activities of these molecules. PIP₂ is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP₂ to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP₃). Once formed IP₃ can diffuse to the endoplasmic reticulum surface where it can bind an IP₃ receptor, e.g., a calcium channel protein containing an IP₃ binding site. IP₃ binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP₃ can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP₄), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. IP₃ and IP₄ can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-bisphosphate (IP₂) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP₂. The other second messenger produced by the hydrolysis of PIP₂, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP₂ or one of its metabolites.

Another signaling pathway in which a receptor protein of the invention may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of

the enzyme adenylyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

Polypeptides

The invention is based on the identification of novel seven-transmembrane proteins/G-coupled protein receptors. Specifically, an expressed sequence tag (EST) was selected based on homology to G-protein-coupled receptor sequences or motifs (e.g., seven-transmembrane domains). This EST was used to design primers based on sequences that it contains and used to identify the 17724, 31945, and 50288 cDNAs. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the assembled sequences revealed that the cloned cDNA molecules encode a G-protein coupled receptor (17724) or putative G-protein coupled receptors (31945 and 50288).

The invention thus relates to a novel GPCR having the deduced amino acid sequence shown in SEQ ID NO:2 or having the amino acid sequence encoded by the deposited cDNA, ATCC No. _____.

The invention also thus relates to a novel putative GPCR having the deduced amino acid sequence shown in SEQ ID NO:5 or having the amino acid sequence encoded by the deposited cDNA, ATCC No. _____.

The invention also thus relates to a novel putative GPCR having the deduced amino acid sequence shown in SEQ ID NO:8 or having the amino acid sequence encoded by the deposited cDNA, ATCC No. _____.

The deposits were made with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia, on _____ and will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms. The deposits are provided as a convenience to those of skill in the art and

is not an admission that a deposit is required under 35 U.S.C. §112. The deposited sequences, as well as the polypeptides encoded by the sequences, are incorporated herein by reference and control in the event of any conflict, such as a sequencing error, with description in this application.

5 The "17724 polypeptide" or "17724 protein" refers to the polypeptide in SEQ ID NO:2 or encoded by the deposited cDNA. The "31945 polypeptide" or "31945 protein" refers to the polypeptide in SEQ ID NO:5 or encoded by the deposited cDNA. The "50288 polypeptide" or "50288 protein" refers to the polypeptide in SEQ ID NO:8 or encoded by the deposited cDNA. The term "protein" or "polypeptide", however, further includes the
10 numerous variants of 17724, 31945, and 50288 polypeptides described herein, as well as fragments derived from the full length 17724, 31945, and 50288 polypeptides and variants.

The present invention thus provides isolated or purified 17724, 31945, and 50288 polypeptides and variants and fragments thereof.

15 The 17724 polypeptide is a 399 residue protein with predicted transmembrane segments as described further in Example 1. PFAM analysis shows homology between amino acid residues 125-374 of SEQ ID NO:2 with a seven transmembrane receptor of the rhodopsin family. A sequence corresponding to the GPCR signature (DRY) is found in the sequence DRF, containing the invariant arginine and amino acids 205-207 shown in SEQ ID NO:2.

20 The 31945 polypeptide is a 663 residue protein. PSORT prediction of protein localization shows a high probability of being found in the endoplasmic reticulum and some probability of being found in vesicles of the secretory system. Putative transmembrane segments are described in further detail in Example 2. PFAM analysis shows homology to a zinc finger, C3HC4 type (ring finger) domain from about amino
25 acids 537 to 574 of SEQ ID NO:5 and to a STAT domain from about amino acids 219 to 225 of SEQ ID NO:5. HMM analysis further shows homology with ring-2 domains from about amino acids 537-574 of SEQ ID NO:5. The highest matches using Prodom analysis are to the human TRC8, which is a multiple membrane spanning receptor. Specifically, amino acids from about 176 to 536 of SEQ ID NO:5 share 27% identity to
30 the ProDom consensus sequence found in polypeptides of TRC8-related protein.

The 50288 polypeptide is a 372 residue protein. PSORT prediction of protein localization shows a high probability of being associated in the nucleus with significant probability of being associated with mitochondria and some probability of being cytoplasmic or extracellular, including cell wall. PFAM analysis shows homology with
5 TNFR/NGFR cysteine-rich region from about amino acids 278 to 308 of SEQ ID NO:8. ProDom analysis show matches to the LYMST/cysteine-rich neurotrophic factor precursor signal. Specifically, amino acids from about amino acids 70 to 113 have approximately 43% sequence identity to this ProDom consensus sequence.

In one embodiment, a polypeptide of the invention includes at least one
10 transmembrane domain. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 15 amino acid residues in length that spans a phospholipid membrane. More preferably, a transmembrane domain includes about at least 18, 20, 22, 24, 25, 30, 35 or 40 amino acid residues and spans a phospholipid membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an α -
15 helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, <http://pfam.wustl.edu/cgi-bin/getdesc?name=7tm-1>, and Zagotta W.N. *et al.* (1996) *Annual Rev. Neurosci.* 19:235-63, the contents of which are incorporated herein
20 by reference.

In a preferred embodiment, a polypeptide of the invention has at least one transmembrane domain or a region which includes at least 18, 20, 22, 24, 25, 30, 35 or 40 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% sequence identity with a "transmembrane domain," e.g., at least one transmembrane
25 domain of human 17724 or 31945.

In another embodiment, a 17724 or 31945 protein includes at least one "non-transmembrane domain." As used herein, "non-transmembrane domains" are domains that reside outside of the membrane. When referring to plasma membranes, non-transmembrane domains include extracellular domains (i.e., outside of the cell) and
30 intracellular domains (i.e., within the cell). When referring to membrane-bound proteins

found in intracellular organelles (e.g., mitochondria, endoplasmic reticulum, peroxisomes and microsomes), non-transmembrane domains include those domains of the protein that reside in the cytosol (i.e., the cytoplasm), the lumen of the organelle, or the matrix or the intermembrane space (the latter two relate specifically to mitochondria organelles). The C-terminal amino acid residue of a non-transmembrane domain is adjacent to an N-terminal amino acid residue of a transmembrane domain in a naturally occurring 17724 or 31945 polypeptide.

In a preferred embodiment, a 17724 or 31945 polypeptide or protein has a "non-transmembrane domain" or a region which includes at least about 1-350, preferably about 200-320, more preferably about 230-300, and even more preferably about 240-280 amino acid residues, and has at least about 60%, 70% 80% 90% 95%, 99% or 100% sequence identity with a "non-transmembrane domain", e.g., a non-transmembrane domain of human 17724 or 31945. Preferably, a non-transmembrane domain is capable of catalytic activity (e.g., modulating signal transduction or ligand bind).

A non-transmembrane domain located at the N-terminus of a 17724 or 31945 protein or polypeptide is referred to herein as an "N-terminal non-transmembrane domain." As used herein, an "N-terminal non-transmembrane domain" includes an amino acid sequence having about 1-350, preferably about 30-325, more preferably about 50-320, or even more preferably about 80-310 amino acid residues in length and is located outside the boundaries of a membrane. For example, an N-terminal non-transmembrane domain is located at about amino acid residues 1-9 of SEQ ID NO:2, or about amino acids 1-8 of SEQ ID NO:5.

Similarly, a non-transmembrane domain located at the C-terminus of a 17724 or 31945 protein or polypeptide is referred to herein as a "C-terminal non-transmembrane domain." As used herein, an "C-terminal non-transmembrane domain" includes an amino acid sequence having about 1-300, preferably about 15-290, preferably about 20-270, more preferably about 25-255 amino acid residues in length and is located outside the boundaries of a membrane. For example, an C-terminal non-transmembrane domain is located at about amino acid residues 378 to 399 of SEQ ID NO:2 or about amino acids 601-663 of SEQ ID NO:5.

5 A 17724, 31945, or 50288 molecule can further include a signal sequence. As used herein, a "signal sequence" refers to a peptide of about 20-80 amino acid residues in length which occurs at the N-terminus of secretory and integral membrane proteins and which contains a majority of hydrophobic amino acid residues. For example, a signal sequence contains at least about 12-25 amino acid residues, preferably about 30-70 amino acid residues, more preferably about 61 amino acid residues, and has at least about 40-70%, preferably about 50-65%, and more preferably about 55-60% hydrophobic amino acid residues (e.g., alanine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, or proline). Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer. For example, in one embodiment, a 17724 or 50288 or 31945 protein contains a signal sequence of about amino acids 1-50 of SEQ ID NO:2, about amino acids 1-42 of SEQ ID NO:8, or about amino acids 1-34 of SEQ ID NO:5, respectfully. The "signal sequence" is cleaved during processing of the mature protein. The mature 17724, 50288 or 31945 proteins correspond to amino acids 50 to 399 of SEQ ID NO:2, amino acids 43 to 372 of SEQ ID NO:8, and amino acids 35 to 663 of SEQ ID NO:5, respectfully.

As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

In one embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other

proteins, or less than about 5% other proteins. When the polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

5 A polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

 The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially
10 free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

 In one embodiment, the 17724, 31945, or 50288 polypeptide comprises the amino
15 acid sequence shown in SEQ ID NOS:2, 5 and 8. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial
20 homology to the 17724, 31945, or 50288 protein of SEQ ID NOS:2, 5 and 8. Variants also include proteins substantially homologous to the 17724, 31945, or 50288 protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the 17724, 31945, or 50288 protein that are produced by
25 chemical synthesis. Variants also include proteins that are substantially homologous to the 17724, 31945, or 50288 protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

 As used herein, two proteins (or a region of the proteins) are substantially
homologous to the 17724, 31945, or 50288 protein when the amino acid sequences are at
least about 40-45%, 45-50%, 50-55%, 55-60%, typically at least about 60-65%, 65-70%, or
30 70-75%, more typically at least about 70-75%, 75-80%, or 80-85%, and most typically at

least about 85-90% or 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NOS:1, 3, 4, 6, 7, and 9 under stringent conditions as more fully described

5 below.

By "variants" is intended proteins or polypeptides having an amino acid sequence that is at least about 60%, 65%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identical to the amino acid sequence of SEQ ID NOS:2, 5, or 8. Variants also include polypeptides encoded by the cDNA insert of the plasmid deposited with ATCC No. _____, _____, and
10 _____, or polypeptides encoded by a nucleic acid molecule that hybridizes to the nucleic acid molecule of SEQ ID NOS:1, 3, 4, 6, 7 or 9, or a complement thereof, under stringent conditions. In another embodiment, a variant of an isolated polypeptide of the present invention differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues from the sequence shown in SEQ ID NOS:2, 5 and 8. If alignment is needed for this comparison
15 the sequences should be aligned for maximum identity. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences. Such variants generally retain the functional activity of the proteins of the invention. Variants include polypeptides that differ in amino acid sequence due to natural allelic variation or mutagenesis.

The invention also encompasses polypeptides having a lower degree of identity but
20 having sufficient similarity so as to perform one or more of the same functions performed by the 17724, 31945, or 50288 polypeptides. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one
25 for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science*
30 247:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%,

even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid “identity” is equivalent to amino acid or nucleic acid “homology”). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished by well-known methods such as using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., Ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., Ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part 1, Griffin, A.M., and Griffin, H.G., Eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., Eds., M Stockton Press, New York, 1991).

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid

"identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

5 The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) *J. Mol. Biol.* 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at
10 <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at
15 <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) is using a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

20 The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (1989) *CABIOS* 4:11-17 which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

 The nucleic acid and protein sequences described herein can be used as a "query
25 sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid
30 molecules of the invention. BLAST protein searches can be performed with the

XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped
5 BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these.

10 Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variants can retain the function of one or more of the regions corresponding to, for example, ligand binding, membrane association, G-protein binding and signal transduction.

15 Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

20 Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

25 As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for the polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

30 Useful variations further can include alteration of ligand binding characteristics. For example, one embodiment involves a variation at the binding site that results in binding but not release, or slower release, of ligand. A further useful variation at the same sites can result in a higher affinity for ligand. Useful variations also include changes that provide for affinity for another ligand. Another useful variation can include one that allows binding but

which prevents activation by the ligand. Another useful variation includes variation in the transmembrane G-protein-binding/signal transduction domain that provides for reduced or increased binding by the appropriate G-protein or for binding by a different G-protein than the one with which the receptor is normally associated. Another useful variation provides a fusion protein in which one or more domains or subregions is operationally fused to one or more domains or subregions from another G-protein coupled receptor or other seven transmembrane protein. Further useful variations include variation in GTP binding sites/domains.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

Substantial homology can be to the entire nucleic acid or amino acid sequence or to fragments of these sequences.

The invention thus also includes polypeptide fragments of the 17724, 31945, and 50288 protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NOS:2, 5, or 8. However, the invention also encompasses fragments of the variants of the 17724, 31945, and 50288 protein as described herein.

The fragments *per se* to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention (known fragments are encompassed in uses and methods specific for tissues or disorders with which the gene is associated).

Fragments can retain one or more of the biological activities of the protein, for example, the ability to bind to a G-protein or ligand, as well as fragments that can be used as an immunogen to generate antibodies.

Biologically active fragments of the 17724, 31945, and 50288 protein (peptides which are, for example, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-100, or more amino acids in length) can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments or parts thereof, G-protein binding site, GPCR signature, glycosylation site or phosphorylation site, or any of the other functional sites including, but not limited to, those shown in the figures herein.

Such domains or motifs can be identified by means of routine computerized homology searching procedures.

Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain or parts thereof; 4) any of the specific transmembrane segments, or parts thereof; 5) any of the three intracellular or three extracellular loops, or parts thereof. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments include the mature protein from about amino acid 6 to the last amino acid. Other fragments contain the various functional sites described herein, such as phosphorylation sites, glycosylation sites, cAMP- and cGMP-dependent, protein kinase C, tyrosine kinase, and casein kinase II phosphorylation sites, N-myristoylation sites, glycosaminoglycan attachment sites, immunoglobulin and major histocompatibility complex protein signature, fragments defining membrane association, and a sequence containing the GPCR signature sequence. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived. In no way however are such fragments to be construed as encompassing fragments that may be found in the art. However, it is understood that with regard to uses and methods of the invention, fragments

that may be known prior to the invention are encompassed. These fragments and others may be encompassed in specific methods and uses relating to tissues/disorders in which expression of the genes of the invention is relevant.

These regions can be identified by well-known methods involving computerized
5 homology analysis.

Fragments also include antigenic fragments and specifically in regions shown to have a high antigenic index in Figures 2, 10 and 13.

Accordingly, possible fragments include, but are not limited to, fragments defining a
ligand-binding site, fragments defining a glycosylation site, fragments defining membrane
10 association, fragments defining phosphorylation sites, and fragments defining interaction
with G proteins and signal transduction, and any of the other functional activities such as
those shown in the figures herein. By this is intended a discrete fragment that provides the
relevant function or allows the relevant function to be identified. In a preferred
embodiment, the fragment contains the ligand-binding site.

15 The invention also provides 17724, 31945, and 50288 protein fragments with
immunogenic properties. These contain an epitope-bearing portion of the 17724, 31945,
and 50288 protein and variants. These peptides can contain at least 5-10, 11, 12, 13, at least
14, or between at least about 15 to about 30 amino acids.

Non-limiting examples of antigenic polypeptides that can be used to generate
20 antibodies include peptides derived from the amino terminal extracellular domain or any of
the extracellular loops. Regions having a high antigenicity index are shown in Figures 2, 10,
and 13.

The epitope-bearing receptor and polypeptides may be produced by any
conventional means (Houghten, R.A., *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985)).
25 Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be
within a larger polypeptide. Further, several fragments can be comprised within a single
larger polypeptide. In one embodiment a fragment designed for expression in a host can
have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the
30 fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion proteins. These comprise a protein of the invention operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the protein. "Operatively linked" indicates that the protein of the invention and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the protein of the invention.

In one embodiment the fusion protein does not affect protein function *per se*. For example, the fusion protein can be a GST-fusion protein in which the sequences of the invention are fused to the C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL-4 fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant protein of the invention. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its C- or N-terminus.

EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.* (*J. Mol. Recog.* 8:52-58 (1995)) and Johanson *et al.* (*J. Biol. Chem.* 270, 16:9459-9471 (1995)). Thus, this invention also encompasses soluble fusion proteins containing a polypeptide of the invention and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence which is also incorporated and can be cleaved with factor Xa.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein.

Another form of fusion protein is one that directly affects protein functions. Accordingly, a polypeptide is encompassed by the present invention in which one or more of the receptor domains (or parts thereof) has been replaced by homologous domains (or parts thereof) from another seven-transmembrane protein, for example another G-protein coupled receptor or other type of receptor. Accordingly, various permutations are possible. The amino terminal extracellular domain, or subregion thereof, (for example, ligand-binding) can be replaced with the domain or subregion from another ligand-binding receptor protein. Alternatively, the entire transmembrane domain, or any of the seven segments or loops, or parts thereof, for example, G-protein-binding/signal transduction, can be replaced. Finally, the carboxy terminal intracellular domain or subregion can be replaced. Thus, chimeric seven-transmembrane proteins/receptors can be formed in which one or more of the native domains or subregions has been replaced.

The isolated 17724, 31945, and 50288 protein can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

For example, the 17724 protein can be purified from the cells shown in Figures 4-8 and 14-17 as especially from tumorous lung, esophagus, lymph node, ovary, thyroid, heart, spinal cord, brain, brain cortex, brain hypothalamus, prostate, spleen, cervix and aorta,

kidney, and muscle tissues in which the gene is highly expressed. The isolated 31945 protein can be purified from the tissue shown in Figure 11, and particularly those in which the gene is relatively highly expressed. The 50288 can be purified from tissues that include, but are not limited to, adrenal gland, brain, breast, colon to liver metastases, pituitary, prostate, T-cells and malignant colon.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation,

glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

5 Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this
10 subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* *Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

 As is also well known, polypeptides are not always entirely linear. For instance,
15 polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

20 Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-
25 formylmethionine.

 The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be
30 expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the

same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

The same type of modification may be present in the same or varying degree at
5 several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

Polypeptide uses

The polypeptides are useful for various biological assays as described in detail
10 below. Since the 17724, 31945, or 50288 gene is expressed in the tissues shown in Figures 4-8, 11, and 14-17 or otherwise disclosed herein, the assays are particularly useful in cells derived from these tissue types, and particularly the tissues in which the gene is highly expressed. Furthermore, since the gene is expressed in these tissues, assays involving the protein in pathological tissue/disorders, particularly applies to disorders involving these
15 tissues and especially the tissues in which the gene is highly expressed. The assays and methods involving pathology/disorders are particularly relevant to cardiovascular disease and tissue fibrosis, especially liver fibrosis, and especially where the fibrosis is the result of viral infection. The assays and methods involving pathology/disorders are also particularly relevant in carcinogenesis, especially in the tissues in which the gene is expressed as
20 disclosed herein and, more particularly, in which the gene is highly expressed. The assays and methods involving pathology/disorders are also particularly relevant in disorders involving inflammation/immunology, where gene expression is found or differential expression is found in B or T-cells. Further, the assays and methods involving pathology/disorders are also particularly relevant in disorders involving viral infection.
25 Furthermore, for the 17724 sequence, the assays and methods involving pathology/disorders are particularly relevant in disorders involving pain.

Disorders involving the spleen include, but are not limited to, splenomegaly, including nonspecific acute splenitis, congestive splenomegaly, and splenic infarcts; neoplasms, congenital anomalies, and rupture. Disorders associated with splenomegaly
30 include infections, such as nonspecific splenitis, infectious mononucleosis, tuberculosis,

typhoid fever, brucellosis, cytomegalovirus, syphilis, malaria, histoplasmosis, toxoplasmosis, kala-azar, trypanosomiasis, schistosomiasis, leishmaniasis, and echinococcosis; congestive states related to partial hypertension, such as cirrhosis of the liver, portal or splenic vein thrombosis, and cardiac failure; lymphohematogenous disorders, such as Hodgkin disease, non-Hodgkin lymphomas/leukemia, multiple myeloma, myeloproliferative disorders, hemolytic anemias, and thrombocytopenic purpura; immunologic-inflammatory conditions, such as rheumatoid arthritis and systemic lupus erythematosus; storage diseases such as Gaucher disease, Niemann-Pick disease, and mucopolysaccharidoses; and other conditions, such as amyloidosis, primary neoplasms and cysts, and secondary neoplasms.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders involving the colon include, but are not limited to, congenital anomalies, such as atresia and stenosis, Meckel diverticulum, congenital aganglionic megacolon-Hirschsprung disease; enterocolitis, such as diarrhea and dysentery, infectious enterocolitis, including viral gastroenteritis, bacterial enterocolitis, necrotizing enterocolitis, antibiotic-associated colitis (pseudomembranous colitis), and collagenous and lymphocytic colitis, miscellaneous intestinal inflammatory disorders, including parasites and protozoa, acquired immunodeficiency syndrome, transplantation, drug-induced intestinal injury, radiation enterocolitis, neutropenic colitis (typhlitis), and diversion colitis; idiopathic inflammatory bowel disease, such as Crohn disease and ulcerative colitis; tumors of the colon, such as non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease, α_1 -antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrahepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and

tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

Disorders involving the uterus and endometrium include, but are not limited to, endometrial histology in the menstrual cycle; functional endometrial disorders, such as anovulatory cycle, inadequate luteal phase, oral contraceptives and induced endometrial changes, and menopausal and postmenopausal changes; inflammations, such as chronic endometritis; adenomyosis; endometriosis; endometrial polyps; endometrial hyperplasia; malignant tumors, such as carcinoma of the endometrium; mixed Müllerian and mesenchymal tumors, such as malignant mixed Müllerian tumors; tumors of the myometrium, including leiomyomas, leiomyosarcomas, and endometrial stromal tumors.

Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in

children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute

5 necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including

10 striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as

15 leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of

20 metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephalopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma,

25 oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor

30 (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including

neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

Disorders involving T-cells include, but are not limited to, cell-mediated hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated cytotoxicity, and
5 transplant rejection; autoimmune diseases, such as systemic lupus erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies, mixed connective tissue disease, and polyarteritis nodosa and other vasculitides; immunologic deficiency syndromes, including but not limited to, primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory)
10 proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell
15 leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.

Diseases of the skin, include but are not limited to, disorders of pigmentation and melanocytes, including but not limited to, vitiligo, freckle, melasma, lentigo, nevocellular nevus, dysplastic nevi, and malignant melanoma; benign epithelial tumors, including but not limited to, seborrheic keratoses, acanthosis nigricans, fibroepithelial polyp, epithelial cyst,
20 keratoacanthoma, and adnexal (appendage) tumors; premalignant and malignant epidermal tumors, including but not limited to, actinic keratosis, squamous cell carcinoma, basal cell carcinoma, and merkel cell carcinoma; tumors of the dermis, including but not limited to, benign fibrous histiocyoma, dermatofibrosarcoma protuberans, xanthomas, and dermal vascular tumors; tumors of cellular immigrants to the skin, including but not limited to,
25 histiocytosis X, mycosis fungoides (cutaneous T-cell lymphoma), and mastocytosis; disorders of epidermal maturation, including but not limited to, ichthyosis; acute inflammatory dermatoses, including but not limited to, urticaria, acute eczematous dermatitis, and erythema multiforme; chronic inflammatory dermatoses, including but not limited to, psoriasis, lichen planus, and lupus erythematosus; blistering (bullous) diseases,
30 including but not limited to, pemphigus, bullous pemphigoid, dermatitis herpetiformis, and

noninflammatory blistering diseases: epidermolysis bullosa and porphyria; disorders of epidermal appendages, including but not limited to, acne vulgaris; panniculitis, including but not limited to, erythema nodosum and erythema induratum; and infection and infestation, such as verrucae, molluscum contagiosum, impetigo, superficial fungal infections, and
5 arthropod bites, stings, and infestations.

In normal bone marrow, the myelocytic series (polymorphonuclear cells) make up approximately 60% of the cellular elements, and the erythrocytic series, 20-30%.

Lymphocytes, monocytes, reticular cells, plasma cells and megakaryocytes together constitute 10-20%. Lymphocytes make up 5-15% of normal adult marrow. In the bone
10 marrow, cell types are add mixed so that precursors of red blood cells (erythroblasts), macrophages (monoblasts), platelets (megakaryocytes), polymorphonuclear leucocytes (myeloblasts), and lymphocytes (lymphoblasts) can be visible in one microscopic field. In addition, stem cells exist for the different cell lineages, as well as a precursor stem cell for the committed progenitor cells of the different lineages. The various types of cells and
15 stages of each would be known to the person of ordinary skill in the art and are found, for example, on page 42 (Figure 2-8) of *Immunology, Immunopathology and Immunity*, Fifth Edition, Sell *et al.* Simon and Schuster (1996), incorporated by reference for its teaching of cell types found in the bone marrow. According, the invention is directed to disorders arising from these cells. These disorders include but are not limited to the following:
20 diseases involving hematopoietic stem cells; committed lymphoid progenitor cells; lymphoid cells including B and T-cells; committed myeloid progenitors, including monocytes, granulocytes, and megakaryocytes; and committed erythroid progenitors. These include but are not limited to the leukemias, including B-lymphoid leukemias, T-lymphoid leukemias, undifferentiated leukemias; erythroleukemia, megakaryoblastic leukemia,
25 monocytic leukemia with and without differentiation; chronic and acute lymphoblastic leukemia, chronic and acute lymphocytic leukemia, chronic and acute myelogenous leukemia, lymphoma, myelo dysplastic syndrome, chronic and acute myeloid leukemia, myelomonocytic leukemia; chronic and acute myeloblastic leukemia, chronic and acute myelogenous leukemia, chronic and acute promyelocytic leukemia, chronic and acute
30 myelocytic leukemia, hematologic malignancies of monocyte-macrophage lineage, such as

juvenile chronic myelogenous leukemia; secondary AML, antecedent hematological disorder; refractory anemia; aplastic anemia; reactive cutaneous angioendotheliomatosis; fibrosing disorders involving altered expression in dendritic cells, disorders including systemic sclerosis, E-M syndrome, epidemic toxic oil syndrome, eosinophilic fasciitis

5 localized forms of scleroderma, keloid, and fibrosing colonopathy; angiomatoid malignant fibrous histiocytoma; carcinoma, including primary head and neck squamous cell carcinoma; sarcoma, including kaposi's sarcoma; fibroadenoma and phyllodes tumors, including mammary fibroadenoma; stromal tumors; phyllodes tumors, including histiocytoma; erythroblastosis; neurofibromatosis; diseases of the vascular endothelium;

10 demyelinating, particularly in old lesions; gliosis, vasogenic edema, vascular disease, Alzheimer's and Parkinson's disease; T-cell lymphomas; B-cell lymphomas.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction,

15 chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous

20 degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive

25 cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart

30 disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal

defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and
5 atresia, and aortic stenosis and atresia, atheroscleroris and disorders involving cardiac transplantation.

Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies,
10 such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic angitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger
15 disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetie) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava
20 syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic
25 interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

Disorders involving red cells include, but are not limited to, anemias, such as hemolytic anemias, including hereditary spherocytosis, hemolytic disease due to erythrocyte enzyme defects: glucose-6-phosphate dehydrogenase deficiency, sickle cell disease,
30 thalassemia syndromes, paroxysmal nocturnal hemoglobinuria, immunohemolytic anemia,

and hemolytic anemia resulting from trauma to red cells; and anemias of diminished erythropoiesis, including megaloblastic anemias, such as anemias of vitamin B₁₂ deficiency: pernicious anemia, and anemia of folate deficiency, iron deficiency anemia, anemia of chronic disease, aplastic anemia, pure red cell aplasia, and other forms of marrow failure.

5 Disorders involving the thymus include developmental disorders, such as DiGeorge syndrome with thymic hypoplasia or aplasia; thymic cysts; thymic hypoplasia, which involves the appearance of lymphoid follicles within the thymus, creating thymic follicular hyperplasia; and thymomas, including germ cell tumors, lymphomas, Hodgkin disease, and carcinoids. Thymomas can include benign or encapsulated thymoma, and malignant
10 thymoma Type I (invasive thymoma) or Type II, designated thymic carcinoma.

Disorders involving B-cells include, but are not limited to precursor B-cell neoplasms, such as lymphoblastic leukemia/lymphoma. Peripheral B-cell neoplasms include, but are not limited to, chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma, diffuse large B-cell lymphoma, Burkitt lymphoma, plasma cell
15 neoplasms, multiple myeloma, and related entities, lymphoplasmacytic lymphoma (Waldenström macroglobulinemia), mantle cell lymphoma, marginal zone lymphoma (MALToma), and hairy cell leukemia.

Disorders involving the kidney include, but are not limited to, congenital anomalies including, but not limited to, cystic diseases of the kidney, that include but are not limited to,
20 cystic renal dysplasia, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases of renal medulla, which include, but are not limited to, medullary sponge kidney, and nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-associated) cystic disease, such as simple cysts; glomerular diseases including pathologies of glomerular injury that include,
25 but are not limited to, in situ immune complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann nephritis, and antibodies against planted antigens, circulating immune complex nephritis, antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation of alternative complement pathway, epithelial cell injury, and pathologies involving mediators of glomerular injury including cellular and soluble
30 mediators, acute glomerulonephritis, such as acute proliferative (poststreptococcal,

postinfectious) glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic) glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous nephropathy), minimal change disease (lipoid nephrosis), focal segmental glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy (Berger disease), focal proliferative and necrotizing glomerulonephritis (focal glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome and thin membrane disease (benign familial hematuria), chronic glomerulonephritis, glomerular lesions associated with systemic disease, including but not limited to, systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis, diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid glomerulonephritis, and other systemic disorders; diseases affecting tubules and interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including but not limited to, pyelonephritis and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced by drugs and toxins, including but not limited to, acute drug-induced interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to, urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytoma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

Disorders of the breast include, but are not limited to, disorders of development; inflammations, including but not limited to, acute mastitis, periductal mastitis, periductal mastitis (recurrent subareolar abscess, squamous metaplasia of lactiferous ducts), mammary duct ectasia, fat necrosis, granulomatous mastitis, and pathologies associated with silicone breast implants; fibrocystic changes; proliferative breast disease including, but not limited to, epithelial hyperplasia, sclerosing adenosis, and small duct papillomas; tumors including, but not limited to, stromal tumors such as fibroadenoma, phyllodes tumor, and sarcomas, and epithelial tumors such as large duct papilloma; carcinoma of the breast including in situ (noninvasive) carcinoma that includes ductal carcinoma in situ (including Paget's disease) and lobular carcinoma in situ, and invasive (infiltrating) carcinoma including, but not limited to, invasive ductal carcinoma, no special type, invasive lobular carcinoma, medullary carcinoma, colloid (mucinous) carcinoma, tubular carcinoma, and invasive papillary carcinoma, and miscellaneous malignant neoplasms.

Disorders in the male breast include, but are not limited to, gynecomastia and carcinoma.

Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to, gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion, testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumor of sex cord-gonadal stroma including, but not limited to, leydig (interstitial) cell tumors and sertoli cell tumors (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

Disorders involving the prostate include, but are not limited to, inflammations, benign enlargement, for example, nodular hyperplasia (benign prostatic hypertrophy or hyperplasia), and tumors such as carcinoma.

Disorders involving the thyroid include, but are not limited to, hyperthyroidism; hypothyroidism including, but not limited to, cretinism and myxedema; thyroiditis

including, but not limited to, hashimoto thyroiditis, subacute (granulomatous) thyroiditis, and subacute lymphocytic (painless) thyroiditis; Graves disease; diffuse and multinodular goiter including, but not limited to, diffuse nontoxic (simple) goiter and multinodular goiter; neoplasms of the thyroid including, but not limited to, adenomas, other benign tumors, and carcinomas, which include, but are not limited to, papillary carcinoma, follicular carcinoma, medullary carcinoma, and anaplastic carcinoma; and congenital anomalies.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

Disorders involving the pancreas include those of the exocrine pancreas such as congenital anomalies, including but not limited to, ectopic pancreas; pancreatitis, including but not limited to, acute pancreatitis; cysts, including but not limited to, pseudocysts; tumors, including but not limited to, cystic tumors and carcinoma of the pancreas; and disorders of the endocrine pancreas such as, diabetes mellitus; islet cell tumors, including but not limited to, insulinomas, gastrinomas, and other rare islet cell tumors.

Disorders involving the small intestine include the malabsorption syndromes such as, celiac sprue, tropical sprue (postinfectious sprue), whipple disease, disaccharidase (lactase) deficiency, abetalipoproteinemia, and tumors of the small intestine including adenomas and adenocarcinoma.

Disorders related to reduced platelet number, thrombocytopenia, include idiopathic thrombocytopenic purpura, including acute idiopathic thrombocytopenic purpura, drug-induced thrombocytopenia, HIV-associated thrombocytopenia, and thrombotic microangiopathies: thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome.

Disorders involving precursor T-cell neoplasms include precursor T lymphoblastic leukemia/lymphoma. Disorders involving peripheral T-cell and natural killer cell neoplasms include T-cell chronic lymphocytic leukemia, large granular lymphocytic leukemia, mycosis fungoides and Sézary syndrome, peripheral T-cell lymphoma, unspecified, angioimmunoblastic T-cell lymphoma, angiocentric lymphoma (NK/T-cell lymphoma^{4a}), intestinal T-cell lymphoma, adult T-cell leukemia/lymphoma, and anaplastic large cell lymphoma.

Disorders involving the ovary include, for example, polycystic ovarian disease, Stein-leventhal syndrome, Pseudomyxoma peritonei and stromal hyperthecosis; ovarian

tumors such as, tumors of coelomic epithelium, serous tumors, mucinous tumors, endometrioid tumors, clear cell adenocarcinoma, cystadenofibroma, brenner tumor, surface epithelial tumors; germ cell tumors such as mature (benign) teratomas, monodermal teratomas, immature malignant teratomas, dysgerminoma, endodermal sinus tumor, choriocarcinoma; sex cord-stromal tumors such as, granulosa-theca cell tumors, thecoma-fibromas, androblastomas, Hill cell tumors, and gonadoblastoma; and metastatic tumors such as Krukenberg tumors.

Bone-forming cells include the osteoprogenitor cells, osteoblasts, and osteocytes. The disorders of the bone are complex because they may have an impact on the skeleton during any of its stages of development. Hence, the disorders may have variable manifestations and may involve one, multiple or all bones of the body. Such disorders include, congenital malformations, achondroplasia and thanatophoric dwarfism, diseases associated with abnormal matrix such as type 1 collagen disease, osteoporosis, Paget's disease, rickets, osteomalacia, high-turnover osteodystrophy, low-turnover of aplastic disease, osteonecrosis, pyogenic osteomyelitis, tuberculous osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondromas, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defects, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing's sarcoma, primitive neuroectodermal tumor, giant cell tumor, and metastatic tumors.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary

hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders of the spinal cord include, but are not limited to, spinal cord compression (i.e., tumors of the cord, epidural abscess, epidural hemorrhage and hematomyelia, acute disk protrusion); noncompressive neoplastic myelopathies (i.e., intramedullary metastasis, paracarcinomatous myelopathy and radiation myelopathy); inflammatory myelopathies (i.e., acute myelitis, transverse myelitis, and necrotic myelopathy); spinal cord infarction; vascular malformation of the spinal cord; and chronic myelopathies (i.e., spondylosis, degenerative and inherited myelopathies, subacute combined degeneration due to vitamin B₁₂ deficiency, syringomyelia, and tabes dorsalis).

Further disorder of interest for the 17724 sequence includes pain disorders. Such pain disorders include, but are not limited to, chronic and acute pain, chest discomfort and palpitation, abdominal pain, headache (i.e., migraine, cluster headache, tension headache, etc.), back and neck pain, and neck and shoulder pain. A more complete description of the disorders resulting in such pain conditions can be found in, for example, Isselbacher *et al.* (1994) Harrison's Principles of Internal Medicine (McGraw-Hill, New York) pp. 49-81, herein incorporated by reference.

The polypeptides of the invention are useful for producing antibodies specific for the 17724, 31945, or 50288 protein, regions, or fragments. Regions having a high antigenicity index score are shown in Figures 2, 10, and 13.

The polypeptides, variants, and fragments (including those which may have been disclosed prior to the present invention) are useful for biological assays related to seven-transmembrane proteins/GPCRs. Such assays involve any of the known seven-

transmembrane protein/GPCR functions or activities or properties useful for diagnosis and treatment of seven-transmembrane protein/GPCR-related conditions.

The polypeptides of the invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally
5 express the protein, as a biopsy or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing the protein.

Determining the ability of the test compound to interact with the polypeptide can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of the ligand, or a biologically active portion thereof,
10 to bind to the polypeptide.

The polypeptides can be used to identify compounds that modulate protein activity. Such compounds, for example, can increase or decrease affinity or rate of binding to a known ligand, compete with ligand for binding to the protein, or displace ligand or substrate bound to the protein. The 17724, 31945, and 50288 protein and appropriate variants and
15 fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the protein. These compounds can be further screened against a functional protein to determine the effect of the compound on the protein activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the protein to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or,
20 alternatively, *in vivo* (e.g., by administering the agent to a subject). Examples for the 17724, 31945, and 50288 protein include but are not limited to purine analogs such as those discussed above. Examples for the 38911 protein include but are not limited to C5a and C5a analogs.

The polypeptides of the invention can be used to screen a compound for the ability
25 to stimulate or inhibit interaction between the protein and a target molecule that normally interacts with the protein. The target can be ligand or a component of the signal pathway with which the protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). The assay includes the steps of combining the protein with a
30 candidate compound under conditions that allow the protein or fragment to interact with the

target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cyclic AMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406);

(Cwiria *et al.* (1990) *Proc. Natl. Acad. Sci.* 97:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra*).

Candidate compounds include, for example, 1) ligand or ligand analogs; 2) peptides such as soluble peptides, Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)), and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 3) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 4) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 5) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble full-length protein or fragment that competes for ligand binding. Other candidate compounds include mutant proteins or appropriate fragments containing mutations that affect protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the protein, or a protein target, could also be measured.

Binding and/or activating compounds can also be screened by using chimeric proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein than that which

is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the protein is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The polypeptides of the invention are also useful in competition binding assays in methods designed to discover compounds that interact with the protein. Thus, a compound is exposed to a polypeptide of the invention under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble polypeptide is also added to the mixture. If the test compound interacts with the soluble polypeptide, it decreases the amount of complex formed or activity from the protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the protein. Thus, the soluble polypeptide that competes with the target region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is desirable to immobilize either the protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/17724,

31945, or 50288 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a protein of the invention-binding protein and a candidate compound are incubated in the protein of the invention-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the protein target molecule, or which are reactive with protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Modulators of 17724, 31945, or 50288 protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, by treating cells that express the 17724, 31945, or 50288 protein, and especially highly express it, such as in the figures disclosed herein or otherwise disclosed herein. These assays are preferably performed in cells related to the disorders as disclosed hereinabove. For example in congestive heart failure, ischemia, and myopathy cells could be cardiomyocytes. Methods of treatment include the steps of administering the modulators of protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

The polypeptides of the invention are thus useful for treating a protein of the invention-associated disorder characterized by aberrant expression or activity of a protein of the invention . In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that
5 modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another embodiment, the method involves administering a protein as therapy to compensate for reduced or aberrant expression or activity of the protein.

Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a
10 beneficial effect. Likewise, inhibition of protein activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example of
15 such a situation, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example of such a situation, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

In yet another aspect of the invention, the proteins of the invention can be used as
20 "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their
25 activity.

The 17724, 31945, and 50288 polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the protein, especially in cells including, but not limited to, those disclosed herein such as in the figures or otherwise disclosed, and especially cells in which the gene is highly expressed. Disorders, thus
30 include diseases of any tissue in which the gene is expressed. Tissue disorders are described

in more detail hereinabove, and particularly relevant disorders are pointed out. Accordingly, methods are provided for detecting the presence, or levels of, the protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the protein such that the interaction can be detected.

5 One agent for detecting a protein of the invention is an antibody capable of selectively binding to the protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

 The proteins of the invention also provide a target for diagnosing active disease, or predisposition to disease, in a patient having a variant protein. Thus, a protein of the
10 invention can be isolated from a biological sample, assayed for the presence of a genetic mutation that results in an aberrant protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered protein activity in cell-based or cell-free assay,
15 alteration in ligand, or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein.

In vitro techniques for detection of the protein include enzyme linked
 immunosorbent assays (ELISAs), Western blots, immunoprecipitations and
20 immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of a protein of the invention expressed in a subject and methods which detect
25 fragments of a protein of the invention in a sample.

 The polypeptides of the invention are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M., *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996), and Linder,
30 M.W., *Clin. Chem.* 43(2):254-266 (1997). The clinical outcomes of these variations result

in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the protein in which one or more of the protein functions in one population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

The polypeptides of the invention are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or activity can be monitored over the course of treatment using the polypeptides as an end-point target. The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of a specified protein in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-

administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The polypeptides of the invention are also useful for treating an associated disorder. Accordingly, methods for treatment include the use of soluble protein or fragments of the protein that compete for ligand binding. These proteins or fragments can have a higher affinity for the ligand so as to provide effective competition.

Treatment is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. "Subject", as used herein, can refer to a mammal, *e.g.* a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, *e.g.* a horse, cow, goat, or other domestic animal. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

Antibodies

The invention also provides antibodies that selectively bind to the 17724, 31945, or 50288 proteins and variants and fragments. An antibody is considered to selectively bind, even if it also binds to other proteins that are not substantially homologous with the proteins. These other proteins share homology with a fragment or domain of the protein of the invention. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to the protein of the invention is still selective.

To generate antibodies, an isolated polypeptide of the invention is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figures 2, 10 and 12.

Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as

described herein. A preferred fragment produces an antibody that diminishes or completely prevents ligand-binding. Antibodies can be developed against the entire protein or portions of the protein, for example, the intracellular carboxy terminal domain, the amino terminal extracellular domain, the entire transmembrane domain or specific segments, any of the intra or extracellular loops, or any portions of the above. Antibodies may also be developed against specific functional sites, such as the site of ligand-binding, the site of G protein coupling, or sites that are phosphorylated, glycosylated, or myristoylated.

An antigenic 17724, 31945, and 50288 fragment will typically comprise at least 8-10 contiguous amino acid residues. The antigenic peptide can comprise, however, a contiguous sequence of at least 12, 14 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, or at least 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as encompassing any fragments which may be disclosed prior to the invention.

Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or F(ab')₂) can be used.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ¹²⁵I, ¹³¹I, ³⁵S or ³H.

An appropriate immunogenic preparation can be derived from native, recombinantly expressed, protein or chemically synthesized peptides.

Antibody Uses

The antibodies can be used to isolate a protein of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells.

The antibodies are useful to detect the presence of a protein of the invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development.

The antibodies can be used to detect a protein of the invention *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

Antibody detection of circulating fragments of the full length protein can be used to identify protein turnover.

Further, the antibodies can be used to assess expression of a protein of the invention in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to protein function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of the protein, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole protein or portions of the receptor, for example, portions of the amino terminal extracellular domain or extracellular loops.

The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting protein expression level or the presence of aberrant proteins of the invention and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies
5 prepared against polymorphic proteins of the invention can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

10 The antibodies are also useful for tissue typing. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific
15 polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

The antibodies are also useful for inhibiting protein function, for example, blocking ligand binding.

These uses can also be applied in a therapeutic context in which treatment involves
20 inhibiting protein function. An antibody can be used, for example, to block ligand binding. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein associated with a cell.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. For an overview of this technology for producing human antibodies, see
25 Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806.

The invention also encompasses kits for using antibodies to detect the presence of a
30 protein of the invention in a biological sample. The kit can comprise antibodies such as a

labeled or labelable antibody and a compound or agent for detecting the protein in a biological sample; means for determining the amount of the protein in the sample; and means for comparing the amount of the protein in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise
5 instructions for using the kit to detect the protein.

Polynucleotides

The nucleotide sequence in SEQ ID NOS:1, 3, 4, 6, 7, and 9 was obtained by sequencing the deposited human full length cDNA. Accordingly, the sequence of the
10 deposited clone is controlling as to any discrepancies between the two and any reference to the sequence of SEQ ID NOS:1, 3, 4, 6, 7, and 9 includes reference to the sequence of the deposited cDNA.

The specifically disclosed cDNAs comprise the coding region and 5' and 3' untranslated sequences (SEQ ID NOS:1, 3, 4, 6, 7, and 9).

15 The nucleotide sequences in SEQ ID NOS:1, 3, 4, 6, 7, and 9 encode full length proteins corresponding to those described in SEQ ID NOS:2, 5, and 8. Nucleic acid expression includes, but is not limited to, that shown in Figures 4-8 and 11 or otherwise disclosed herein.

The invention provides isolated polynucleotides encoding a 17724, 31945, or 50288
20 protein. The term "17724 polynucleotide" or "17724 nucleic acid" refers to the sequence shown in SEQ ID NOS:1, 3, or in the deposited cDNA. The term "31945 polynucleotide" or "31945 nucleic acid" refers to the sequence shown in SEQ ID NOS:4, 6, or the deposited cDNA. The term "50288 polynucleotide" or "50288 nucleic acid" refers to the sequence shown in SEQ ID NOS:7, 9, or the deposited cDNA.

25 The term "polynucleotide" or "nucleic acid" further includes variants and fragments of the 17724, 31945, and 50288 polynucleotides.

An "isolated" nucleic acid of the invention is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5'
30 and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic

acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences of the invention.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a molar basis) of all macromolecular species present.

The polynucleotides of the invention can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

The polynucleotides of the invention include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Polynucleotides of the invention can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

One nucleic acid comprises a nucleotide sequence shown in SEQ ID NOS:1, 3, 4, 6, 7, and 9, corresponding to human 17724, 31945, and 50288 cDNA.

In one embodiment, the nucleic acid comprises only the coding region. The invention further provides variant polynucleotides, and fragments thereof, that differ from a nucleotide sequence shown in SEQ ID NOS:1, 3, 4, 6, 7, and 9 due to degeneracy of the genetic code and thus encode the same protein as that encoded by a nucleotide sequence shown in SEQ ID NOS:1, 3, 4, 6, 7, and 9.

The invention also provides nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Typically, variants have a substantial identity with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, 6, 7, and 9 and the complements thereof.

Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. Generally, the nucleotide sequence variants of the invention will have at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the nucleotide sequence disclosed herein or fragments thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to a nucleotide sequence shown in SEQ ID NOS:1, 3, 4, 6, 7, and 9 or a fragment of the sequence.

It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all seven-transmembrane proteins, all GPCRs or all family I GPCRs. Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology* John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred, example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization

limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NOS:1, 3, 4, 6, 7, and 9 corresponds to a naturally-occurring nucleic acid molecule.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In one embodiment, an isolated nucleic acid molecule that hybridizes under stringent conditions to a sequence of SEQ ID NOS:1, 3, 4, 6, 7, and 9 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). Timing of hybridization can vary from ½ hour to 10 hours or longer. Shorter hybridizations however can include from 1 to 5, and 6 to 10 hours. Typically, hybridization is performed overnight for around 10-12 hours. The time of washes can also vary from around 10 minutes to 30 minutes. Typically, washes are performed from 10-20 minutes.

As understood by those of ordinary skill, the exact conditions can be determined empirically and depend on ionic strength, temperature and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS. Other factors considered in determining the desired hybridization conditions include the length of the nucleic acid sequences, base composition, percent mismatch between the hybridizing sequences and the frequency of occurrence of subsets of the sequences within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules.

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 4, 6, 7, and 9 and the complements of SEQ ID NOS:1, 3, 4, 6, 7, and 9. In one embodiment, the nucleic

acid consists of a portion of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 4, 6, 7, and 9 and the complements SEQ ID NOS:1, 3, 4, 6, 7, and 9.

The isolated fragments can be at least between 5-10, 10-20, 20-30, 30-40, 40-50, including but not limited to 50, 75, 100, 200, 250, 500, 600, 700, 800, 1000, 1200, 1400, 1500

5 nucleotides in length or greater. Alternatively, a nucleic acid molecule that is a fragment of a sequence of the invention comprises a nucleotide sequence consisting of nucleotides 1-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1100, 1100-1200, 1200-1300, 1300-1400, 1400-1500, 1500-1600, 1600-1700, 1700-1800 of SEQ ID NOS:1, 3, 4, 6, 7, and 9. Fragments which encode antigenic proteins or
10 polypeptides described herein are useful. The fragment can be single or double stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

Other fragments of all four proteins include nucleotide sequences encoding the amino acid fragments described herein. Further, fragments can include subfragments of the
15 specific domains or sites described herein. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention except as they are used in methods involving tissues/disorders with which gene expression is
20 associated.

However, it is understood that a nucleic acid fragment includes any nucleic acid sequence that does not include the entire gene.

Nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. Nucleic acid
25 fragments include but are not limited to nucleic acid molecules encoding a polypeptide comprising an amino terminal extracellular domain, comprising a region spanning the transmembrane domain, a polypeptide comprising a carboxy terminal intracellular domain, and a polypeptide encoding a G-protein receptor signature (the three amino acids or surrounding amino acid residues from about 10 before to about 10 after), nucleic acid
30 molecules encoding any of the seven transmembrane segments, extracellular or intracellular

loops, glycosylation sites, protein kinase C, cAMP, cGMP, and casein kinase II phosphorylation sites, myristoylation sites, glycosaminoglycan attachment site and immunoglobulin and major histocompatibility complex protein signature site, or any other functional sites contained in the proteins, including, not but limited to those disclosed in the figures herein.

Where the location of the domains have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

Nucleic acid fragments also include combinations of the domains, segments, loops, and other functional sites described above. Thus, for example, a nucleic acid could include sequences corresponding to the amino terminal extracellular domain and one transmembrane segment. A person of ordinary skill in the art would be aware of the many permutations that are possible.

Where the location of the domains or sites have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

The invention also provides nucleic acid fragments that encode epitope bearing regions of the proteins described herein.

The isolated polynucleotide sequences, and especially fragments, are useful as DNA probes and primers.

For example, the coding region of a gene of the invention can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of the genes of the invention.

A probe/primer typically comprises substantially purified oligonucleotide. The 17724, 31945, and 50288 oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 10, 20, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NOS:1, 3, 4, 6, 7,

and 9, coding or non-coding, sense or anti-sense strand or other receptor polynucleotides, that hybridize under stringent conditions.

Polynucleotide Uses

5 The nucleic acid sequences of the present invention can be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. As discussed in more detail above, such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the
10 NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST)
15 can be used. See <http://www.ncbi.nlm.nih.gov>.

 The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. “Probes” are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science* 254:1497-1500. Typically, a probe
20 comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of a nucleic acid selected from the group consisting of SEQ ID NOS:1, 3, 4, 6, 7, 9 and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

25 As used herein, the term “primer” refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term “primer site” refers to the area of the target DNA to which a primer
30 hybridizes. The term “primer pair” refers to a set of primers including a 5' (upstream)

primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

The polynucleotides are useful for various biological assays as described in detail below. As disclosed herein, the genes are expressed in various tissues, for example, as shown in the figures or otherwise disclosed herein. Accordingly, the assays are particularly useful in cells derived from these tissue types, and particularly the tissues in which the gene is highly expressed, such as are disclosed in the figures herein or otherwise. Furthermore, since the gene is expressed in these tissues, assays involving the polynucleotides in pathological tissue/disorders, particularly applies to disorders involving these tissues and especially the tissues in which the gene is highly expressed. Disorders in which the genes are particularly relevant and to which the assays particularly apply have been disclosed hereinabove with reference to the section disclosing polypeptide uses, for example, cardiovascular disease and disorders involving pain. As one further example, where a gene is expressed in hemapoeietic precursor cells, the assays and methods involving pathology/disorders related to immune dysfunction and inflammation are particularly relevant. Further, where a gene is expressed in viral infection, assays and methods involving pathology/disorders are particularly relevant in this type of disorder. Further, where a gene is expressed in tissue fibrosis and particularly where a gene is expressed in liver fibrosis, the assays and methods involving pathology/disorders are particularly relevant in this disorder. Finally, where a gene is highly expressed in bone-forming precursors, assays and methods involving osteoporosis and osteopetrosis are particularly relevant.

The receptor polynucleotides are useful for probes, primers, and in biological assays.

Where the polynucleotides are used to assess seven-transmembrane protein/GPCR properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. In this case, even fragments that may have been known prior to the invention are encompassed. Thus, for example, assays specifically directed to seven-transmembrane protein/GPCR functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing protein function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving treatment of protein

dysfunction, all fragments are encompassed including those which may have been known in the art.

The 17724, 31945, and 50288 polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NOS:2, 5, and 8 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NOS:2, 5, and 8 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NOS:2, 5, and 8 was isolated, different tissues from the same organism, or from different organisms.

This method is useful for isolating genes and cDNA that are developmentally-controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene encoding the protein. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. Probes, however, are not to be construed as corresponding to any sequences that may be known prior to the invention.

The 17724, 31945, and 50288 nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NOS:1, 3, 4, 6, 7, and 9, respectively, or a fragment thereof, such as an oligonucleotide of at least about 10-15, 15-20, 25-30, 35-40, 45-50, 50-75, 75-100, 100-200, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

The fragments are also useful to synthesize antisense molecules of desired length and sequence. Antisense nucleic acids of the invention can be designed using the nucleotide sequences of SEQ ID NOS:1, 3, 4, 6, 7, and 9, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using

naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4:5). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci.*

USA 93:14670. PNAs can be further modified, e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup
5 (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63, Mag *et al.* (1989) *Nucleic Acids Res.* 17:5973, and Peterser *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5:1119.

The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell proteins *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl.*
10 *Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm Res.* 5:539-549).

15 The polynucleotides of the invention are also useful as primers for PCR to amplify any given region of the polynucleotide.

The polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide
20 sequence, such as into the cellular genome, to alter *in situ* expression of the genes and gene products. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The polynucleotides are also useful for expressing antigenic peptides. Peptide
25 regions having a high antigenicity index are shown in Figures 2, 10, and 13.

The polynucleotides are also useful as probes for determining the chromosomal positions of the polynucleotides of the invention by means of *in situ* hybridization methods, such as FISH (for a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York), and PCR mapping of somatic

cell hybrids. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

The polynucleotide probes are also useful to determine patterns of the presence of the gene encoding the proteins of the invention and their variants with respect to tissue distribution, for example, whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously.

The polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

The polynucleotides are also useful for constructing host cells expressing a part, or all, of the polynucleotides and polypeptides of the invention.

5 The polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the polynucleotides and polypeptides of the invention.

The polynucleotides are also useful for making vectors that express part, or all, of the polypeptides of the invention.

10 The polynucleotides are also useful as hybridization probes for determining the level of nucleic acid expression of the nucleic acid molecules of the invention. Accordingly, the probes can be used to detect the presence of, or to determine levels of, the nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant
15 in cases in which there has been an amplification of the genes of the invention.

Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the genes of the invention, as on extrachromosomal elements or as integrated into chromosomes in which the gene is not normally found, for example as a homogeneously staining region.

20 These uses are relevant for diagnosis of disorders involving an increase or decrease in expression relative to normal, such as in the disorders described herein.

Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant expression or activity of a nucleic acid of the invention, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is
25 detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells,
30 tissue) to determine whether an individual has a disease or disorder, or is at risk of

developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

5 *In vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

 Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a protein of the invention, such as by measuring a level of a protein-encoding
10 nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a gene of the invention has been mutated.

 Nucleic acid expression assays are useful for drug screening to identify compounds that modulate nucleic acid expression (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs) of the nucleic acid molecules of the invention. A cell is contacted
15 with a candidate compound and the expression of mRNA determined. The level of expression of mRNA of the invention in the presence of the candidate compound is compared to the level of expression of the mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder
20 characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression.

 Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in
25 transgenic animals.

 The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the receptor gene. The method typically includes assaying the ability of the compound to modulate the expression of the nucleic acid and thus identifying a compound that can be used to treat a disorder

characterized by undesired nucleic acid expression of the nucleic acid molecules of the invention.

The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

The assay for nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway (such as cyclic AMP or phosphatidylinositol turnover). Further, the expression of genes that are up- or down-regulated in response to the protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of mRNA in the presence of the candidate compound is compared to the level of expression of mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate nucleic acid expression of the nucleic acid molecules of the invention. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or

improperly modified). Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

Alternatively, a modulator of nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the nucleic acid expression.

The polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The polynucleotides of the invention are also useful in diagnostic assays for qualitative changes in the nucleic acid, and particularly in qualitative changes that lead to pathology. The polynucleotides can be used to detect mutations in genes of the invention and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally-occurring genetic mutations in the gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the

mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a protein of the invention.

Mutations in a gene of the invention can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional

amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially
5 useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Alternatively, mutations in a gene of the invention can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

10 Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

15 Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant gene of the invention and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995)
20 *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or
25 RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing
30 a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et*

al., *Nature* 313:495 (1985)). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility
5 (Keen *et al.* (1991) *Trends Genet.* 7:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or
10 thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7:244-255; Kozal *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the
15 sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary
20 to the mutant gene.

The polynucleotides of the invention are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment
25 (pharmacogenomic relationship). In the present case, for example, a mutation in the gene that results in altered affinity for ligand could result in an excessive or decreased drug effect with standard concentrations of ligand that activates the protein. Accordingly, the polynucleotides described herein can be used to assess the mutation content of the gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

5 The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mRNA, or genomic DNA, such that the presence of mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test sample.

10 "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage;
15 a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or
20 extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

 The polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the
25 chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an
30 amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other

strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a
5 single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

10 The polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

15 Furthermore, the gene sequence can be used to provide an alternative technique which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the receptor sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

20 Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The sequences can be used to obtain such identification sequences from individuals and from
25 tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that

individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

The polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (e.g. blood, saliva, or semen). The amplified sequence can then be compared to a standard allowing identification of the origin of the sample.

The polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique.

The polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, the polynucleotides can be used directly to block transcription or translation of the gene sequences by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable expression of the gene of the invention, nucleic acids can be directly used for treatment.

The polynucleotides are thus useful as antisense constructs to control expression of a gene of the invention in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing

transcription and hence production of the protein of the invention. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into protein.

5 Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NOS:1, 4, and 7 which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NOS:1, 4, and 7.

10 Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of nucleic acid of the invention. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired expression of a nucleic acid of the invention. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the protein of the invention, such as ligand binding.

15 The polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in expression of a gene of the invention. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired protein to treat the individual.

20 The invention also encompasses kits for detecting the presence of a nucleic acid of the invention in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting the nucleic acid in a biological sample; means for determining the amount of the nucleic acid in the sample; and means for comparing the amount of the nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect the mRNA or DNA.

Computer Readable Means

The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software

such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the

computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem.* 17:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/host cells

The invention also provides vectors containing the polynucleotides of the invention. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the polynucleotides. When the vector is a nucleic acid molecule, the polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the polynucleotides of the invention.

Alternatively, the vector may integrate into the host cell genome and produce additional copies of the polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the polynucleotides. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the polynucleotides from the vector. Alternatively, a trans-acting factor may

be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

It is understood, however, that in some embodiments, transcription and/or translation of the polynucleotides can occur in a cell-free system.

5 The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

10 In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

15 In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example,
20 in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

25 A variety of expression vectors can be used to express a polynucleotide of the invention. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for
30 prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A*

Laboratory Manual, 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

5 The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

10 The polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

15 The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

20 It is further recognized that the nucleic acid sequences of the invention can be altered to contain codons, which are preferred, or non-preferred, for a particular expression system. For example, the nucleic acid can be one in which at least one altered codon, and preferably at least 10% or 20% of the codons, have been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells. Methods for determining such codon usage are well known in the art.

25 As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can
30 ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not

limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

- 5 Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

- 10 Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the polynucleotide of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

- 15 The polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943 (1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

- 20 The polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

- 25 In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840 (1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

- 30 The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the polynucleotides. The person of ordinary skill in the art would be aware of other

vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd, ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

5 The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is
10 subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

 The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian
15 cells.

 The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation,
20 transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al. (Molecular Cloning: A Laboratory Manual, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).*

 Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the polynucleotides of the
25 invention can be introduced either alone or with other polynucleotides that are not related to the polynucleotides of the invention such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the polynucleotide vector.

 In the case of bacteriophage and viral vectors, these can be introduced into cells as
30 packaged or encapsulated virus by standard procedures for infection and transduction. Viral

vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

5 Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

10 While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

15 Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the polypeptides of the invention or heterologous to these polypeptides.

20 Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

25 It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

30

Uses of vectors and host cells

It is understood that "host cells" and "recombinant host cells" refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because
5 certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the
10 case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing proteins or polypeptides of the invention that can be further purified to produce desired
15 amounts of the protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

Host cells are also useful for conducting cell-based assays involving the protein or fragments. Thus, a recombinant host cell expressing a native protein of the invention is useful to assay for compounds that stimulate or inhibit protein function. This includes
20 ligand binding, gene expression at the level of transcription or translation, G-protein interaction, and components of the signal transduction pathway.

Host cells are also useful for identifying mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant protein
25 (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native protein.

Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous amino terminal extracellular domain (or other binding region). Alternatively, a
30 heterologous region spanning the entire transmembrane domain (or parts thereof) can be

used to assess the effect of a desired amino terminal extracellular domain (or other binding region) on any given host cell. In this embodiment, a region spanning the entire transmembrane domain (or parts thereof) compatible with the specific host cell is used to make the chimeric vector. Alternatively, a heterologous carboxy terminal intracellular, e.g., signal transduction, domain can be introduced into the host cell.

Further, mutant proteins can be designed in which one or more of the various functions is engineered to be increased or decreased (e.g., ligand binding or G-protein binding) and used to augment or replace the native proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant protein of the invention or providing an aberrant protein that provides a therapeutic result. In one embodiment, the cells provide proteins that are abnormally active.

In another embodiment, the cells provide proteins that are abnormally inactive. These proteins can compete with the endogenous proteins in the individual.

In another embodiment, cells expressing proteins that cannot be activated, are introduced into an individual in order to compete with the endogenous proteins for ligand. For example, in the case in which excessive ligand is part of a treatment modality, it may be necessary to inactivate this ligand at a specific point in treatment. Providing cells that compete for the ligand, but which cannot be affected by receptor activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ* alteration of the endogenous polynucleotide sequences in a host cell genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the polynucleotides or sequences proximal or distal to a gene of the invention are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, a protein of the invention can be produced in a cell not normally producing it. Alternatively, increased expression of the protein can be effected in a cell normally producing the protein at a

specific level. Further, expression can be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the protein sequence or can be a homologous sequence with a desired mutation that affects expression.

Alternatively, the entire gene can be deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant proteins. Such mutations could be introduced, for example, into the specific functional regions such as the ligand-binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell* 51:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous receptor gene is selected (see e.g., Li, E. *et al.*, *Cell* 69:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinions in Biotechnology* 2:823-829 and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic

animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a receptor protein and identifying and evaluating modulators of the protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs,
5 cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which the polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the
10 oocyte to develop in a pseudopregnant female foster animal. Any of the nucleotide sequences of the invention can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if
15 not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar
20 methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals
25 carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, receptor activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* receptor function, including ligand interaction, the effect of specific mutant receptors on receptor function and ligand interaction, and the effect of chimeric receptors. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more receptor functions.

5 In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical compositions

15 The nucleic acid molecules of the invention, protein of the invention (particularly fragments such as the amino terminal extracellular domain), modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

20 As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), 25 transmucosal, and rectal administration. Solutions or suspensions used for parenteral, 30

intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

10 Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be
15 sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures
20 thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents,
25 for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

 Sterile injectable solutions can be prepared by incorporating the active compound
30 (e.g., a seven-transmembrane protein/receptor protein or antibody) in the required amount in

an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of
5 sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can
10 be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared
15 using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth
20 or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an
25 aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For
transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and
30 include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid

derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with
5 conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable,
10 biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to
15 viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used
20 herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular
25 therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see
30 e.g., Chen *et al.*, *PNAS* 91:3054-3057 (1994)). The pharmaceutical preparation of the gene

therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which
5 produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about
10 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease
15 or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time
20 per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described
25 herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or
30 inorganic compounds (i.e., including heteroorganic and organometallic compounds) having

a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

Other Embodiments

In another aspect, the invention features, a method of analyzing a plurality of capture probes. The method can be used, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence; contacting the array with a sequence of the invention, preferably purified, nucleic acid, preferably purified, polypeptide, preferably purified, or antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody of the invention.

The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell. The method can include contacting the nucleic acid, polypeptide, or antibody of the invention with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of a sequence of the invention. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being

positionally distinguishable from each other address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express or mis express sequence of the invention or from a cell or subject in which a 31945, 50288, and 17724 mediated response has been elicited, e.g., by contact of the cell with a 31945, 50288, and 17724 nucleic acid or protein, or administration to the cell or subject 31945, 50288, and 17724 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 31945, 50288, and 17724 nucleic acid, polypeptide, or antibody); providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 31945, 50288, and 17724 (or does not express as highly as in the case of the 31945, 50288, and 17724 positive plurality of capture probes) or from a cell or subject which in which a 31945, 50288, and 17724 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 31945, 50288, and 17724 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

The method can be used to detect SNPs, as described above.

In another aspect, the invention features, a method of analyzing 31945, 50288, and 17724, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 31945, 50288, and 17724 nucleic acid or amino acid sequence; comparing the 31945, 50288, and 17724 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 31945, 50288, and 17724.

Preferred databases include GenBank™. The method can include evaluating the sequence identity between a 31945, 50288, and 17724 sequence and a database sequence.

The method can be performed by accessing the database at a second site, e.g., over the internet.

In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 31945, 50288, and 17724. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality are identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with different labels, such that an oligonucleotide that hybridizes to one allele provides a signal that is distinguishable from an oligonucleotide which hybridizes to a second allele.

EXPERIMENTAL

Example 1. Characterization of the 17724 cDNA

Clone 17724 encodes an approximately 1.9 Kb mRNA transcript having the corresponding cDNA set forth in SEQ ID NO:1. This transcript has a 1200 nucleotide open reading frame shown in SEQ ID NO:3 (nucleotides 323 to 1520 of SEQ ID NO:1), which encodes a 399 amino acid protein (SEQ ID NO:2). An analysis of the full-length 17724 polypeptide predicts that about the N-terminal 50 amino acids represent a signal peptide. Transmembrane segments from about amino acids (aa) 10-34, 41-65, 72-88, 115-135, 143-162, 175-199, 288-245, 284-306, 321-344, and 357-377 were predicted by MEMSAT. Transmembrane segments were also predicted from aa 10-30, 66-86, 94-113, 126-150, 179-196, 235-257, 272-295, and 308-328 of the presumed mature peptide sequence. Prosite program analysis was used to predict various sites within the 17724 protein. N-glycosylation sites were predicted at about aa 89-92, 149-152, and 378-381. A protein kinase C phosphorylation site was predicted at about aa 380-382. Casein kinase II phosphorylation sites were predicted at about aa 103-106, 151-154, and 272-275. N-myristoylation sites were predicted at about aa 217-222, 230-235, 255-260, 326-331, 394-399. A leucine zipper pattern was detected at about aa 271-292, and a G-protein coupled receptor signature was detected at about aa 194-210.

As shown in Figure 3, the 17724 protein possesses a 7 transmembrane receptor domain from the rhodopsin family from aa 125-374, as predicted by HMMer, Version 2. For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 and

5 <http://www.psc.edu/general/software/packages/pfam/pfam.html>. The sequence alignment generated using the Clustal W. Version 1.74 (data not shown) indicated that the protein (17724; SEQ ID NO:2) encoded by human 17724 (SEQ ID NO:1 and SEQ ID NO:3) shares sequence identity to the murine olfactory receptor 6 polypeptide (Genbank Accession Number P34986 and Genbank Accession Number AAD13315).

10 The 17724 polypeptide also shares approximately 53% sequence identity from about amino acid 250-329 and approximately 43% sequence identity from about amino acids 330-394 to the Prodom consensus sequence found in polypeptides from the olfactory receptor-like G-protein coupled transmembrane glycoprotein multigene family. The 17724 sequence also shares approximately 30% sequence identity from about amino acid 139 to about 226 to
15 the Prodom consensus sequence found in members of the transmembrane glycoprotein lipoprotein palmitate protein family.

Example 2: Characterization of the 50288 cDNA

Clone 50288 encodes an approximately 1.6 Kb mRNA transcript having the
20 corresponding cDNA set forth in SEQ ID NO:7. This transcript has a 811 nucleotide open reading frame shown in SEQ ID NO:9 (nucleotides 309 to 1,428 of SEQ ID NO:7), which encodes a 372 amino acid protein (SEQ ID NO:8). An analysis of the full-length 50288 polypeptide predicts that the N-terminal 42 amino acids represents a signal peptide. Prosite program analysis was used to predict various sites within the 50288 protein. N-
25 glycosylation sites were predicted at about amino acids (aa) 153-156. Protein kinase C phosphorylation sites were predicted at about aa 11-13, 18-20, 107-109, 156-158, 224-226, 301-303, 332-334, 335-337. Casein kinase II phosphorylation sites were found from about aa 42-45, 59-62, 81-84, 146-149, 168-171, 282-285, 335-338. Tyrosine kinase phosphorylation sites were found from about aa 50-56 and 109-116. N-myristoylation sites

were found from about aa 77-82, 88-93, 152-157, 268-273, 288-293, 328-333, and 361-366. An RGD cell attachment sequence was found from about aa 162 to about aa 164.

In addition, a sequence alignment generated using the Clustal W. Version 1.74 (data not shown) indicated that the protein (50288; SEQ ID NO:8) encoded by human 50288 (SEQ ID NO:7 and 9) share sequence identity at the N-terminus with a putative serine protease from *Helicoverpa armigera* (Genbank Accession No. Y12274).

Example 3: Characterization of the 31945 cDNA

Clone 31945 encodes at approximately 3.6 Kb mRNA transcript having the corresponding cDNA set forth in SEQ ID NOS:4 and 6. This transcript has a 1991 nucleotide open reading frame shown in SEQ ID NO:6 (nucleotides 332 to 2323 of SEQ ID NO:4) which encodes a 663 amino acid protein (SEQ ID NO:5). An analysis of the full-length 31945 polypeptide predicts that the N-terminal 34 amino acids represent a signal peptide. Transmembrane segments from about amino acids (aa) 9-27, 53-70, 77-93, 123-139, 146-165, 174-198, 205-222, 229-247, 259-275, 282-299, 315-331, 339-358, 382-400, 409-433, 450-474, and 482-500 were predicted by MEMSAT. Transmembrane segments were also predicted from about aa 20-37, 44-60, 90-106, 113-132, 141-165, 172-189, 196-214, 226-242, 249-266, 282-298, 306-325, 349-367, 376-400, 417-441, and 449-467 of the presumed mature peptide sequence. Prosite program analysis was used to predict various sites within the 31945 protein. N-glycosylation sites were found from about aa 455-458, and 580-583. Protein kinase C phosphorylation sites were found from about aa 248-250, 457-459, and 652-654. Casein kinase II phosphorylation sites were predicted at about aa 103-106, 179-182, 248-251, 266-269, 356-359, and 652-655. Tyrosine kinase phosphorylation sites were found from about aa 102-109. N-myristoylation sites were found from about aa 279-284, 314-319, 473-478, 598-603, and 625-630.

Example 4: Tissue Distribution of 17724, 31945, and 50288 mRNA

Northern blot hybridizations with various RNA samples are performed under standard conditions and washed under stringent conditions, i.e., 0.2 X SSC at 65°C. A DNA probe corresponding to all or a portion of the 17224, 31945, and 50288 cDNA (SEQ ID

NOS:1, 3, 4, 6, 7, and 9) can be used. The DNA is radioactively labeled with ³²P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from various tissues and cell lines are probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Expression levels of the 17724 and 31945 sequences were determined by quantitative PCR (Taqman® brand quantitative PCR kit, Applied Biosystems) and are shown in Figures 4-8 and 11. The quantitative PCR reactions were performed according to the kit manufacturer's instructions.

TaqMan analysis of 17724 revealed mRNA expression in a number of tissues including, for example, in clinical tumorous lung samples (Figures 4 and 15), normal and diseased human heart tissues (Figure 6), and various other human tissues (Figures 6, 7, 14, 16, and 17). In addition, TaqMan analysis showed 17724 expression in clinical angiogenic samples. Specifically, elevated expression levels of 17724 mRNA was shown in normal brain tissue when compared to tumorous brain tissue (see Figure 5). 17724 was also found to be highly expressed in spinal cord, brain cortex, and brain hypothalamus as shown in Figure 8.

TaqMan analysis of the 31945 sequence revealed expression in a number of tissues as shown in Figure 11. High level of 31945 mRNA expression was found in the following cell types: NHBE (mock), NHBE IL13-1, ThO (24 hours), and mPB CD34⁺. Moderate levels of 31945 expression were found in the following tissues and cell types: lung, kidney, brain, Hep β2 TβF, erythrold, Mega, and neutrophil. Lower levels of expression of the 31945 sequence were found in fetal liver, Hep β2 (mock), liver fibroblast (NDR), Grans (donors), mBM CD34⁺, CD19, and BM-MNC. Additional tissues which show 31945 mRNA expression are shown in Figure 11.

Example 5: Recombinant Expression of 31945, 50288, and 17724 in Bacterial Cells

In this example, 31945, 50288, or 17724 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 31945, 50288, or 17724 is fused to GST and this fusion

polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-31945, -50288, or -17724 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 6: Expression of Recombinant 31945, 50288, and 17724 Protein in COS Cells

To express the 31945, 50288, or 17724 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 31945, 50288, or 17724 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the 31945, 50288, or 17724 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 31945, 50288, or 17724 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 31945, 50288, or 17724 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 31945, 50288, or 17724 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 α , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the 31945-, 50288-, or 17724-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 31945, 50288, or 17724 polypeptide is detected by radiolabelling (³⁵S-methionine or ³⁵S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ³⁵S-methionine (or ³⁵S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the 31945, 50288, or 17724 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 31945, 50288, or 17724 polypeptide is detected by radiolabelling and immunoprecipitation using a 31945, 50288, or 17724 specific monoclonal antibody.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention pertains having the benefit of the teachings

presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.